

Experimental Political Science and the Study of Causality

From
Nature
to the
Lab

Rebecca B. Morton
Kenneth C. Williams

CAMBRIDGE

MORTON &
WILLIAMS

EXPERIMENTAL POLITICAL SCIENCE
AND THE STUDY OF CAUSALITY

Politika

2489

1

Controlling Observables and Unobservables

4.1 Control in Experiments

4.1.1 Controlling Observables in Experiments

We begin our analysis of the Rubin Causal Model (RCM)-based approaches to estimating the effects of a cause with a review of those that work through the control of observable variables that can make it difficult to estimate causal effects. Specifically, using the notation of the previous chapter, there are two types of observable variables that can cause problems for the estimation of the effects of a cause, Z_i and X_i . Recall that Y_i is a function of X_i and T_i is a function of Z_i . That is, X_i represents the other observable variables that affect our dependent variable besides the treatment variable and Z_i represents the set of observable variables that affect the treatment variable. Moreover, these variables may overlap and we define $W_i = Z_i \cup X_i$.

In experiments researchers deal with these observable variables in two ways – through random assignment and through the ability to manipulate these variables as they do with treatment variables. In the next chapter we show how such random assignment sidesteps both observable and unobservable variables that can interfere with measuring the causal effect of the treatment.

But experimenters also can manipulate some of the observable variables that might have an effect on treatments or directly on voting behavior and thereby reduce their effects. For instance, one observable variable that might affect the treatment variable is the mechanism by which a voter learns the information. We can imagine that if the information is told to subjects verbally, the effect might be different than if the subjects read the information or if it is shown to them visually. In a naturally occurring

election without experimental manipulation or in a field experiment in which the researcher cannot control the mechanism of manipulation, this information may reach voters in a variety of ways, affecting the treatment. In a laboratory experiment, and to some extent in a field experiment, a researcher can control the mechanism so that it does not vary across subjects. Or, if the researcher is interested in the effects of different mechanisms as well as information itself, the researcher can randomly assign different mechanisms to the subjects.

An observable variable that might affect subjects' voting behavior independent of treatment could be the language used to describe the candidates in the election and the other aspects of the election environment. In a naturally occurring election, different voters may be exposed to different descriptions of the candidates and other aspects of the environment that affect their voting behavior. In a laboratory, and to some extent in a field experiment, a researcher can control this language and the other aspects of the election environment that have these effects so that they do not vary across subjects. We call the information provided to subjects during an experiment the script. Or a researcher might randomize the language to reduce possible effects as with the mechanism of providing information. In this way experimentalists can control for W_i . Guala (2005, p. 238) remarks: "[T]he experimental method works by eliminating possible sources of error or, in other words, by controlling systematically the background factors that may induce us to draw a mistaken inference from the evidence to the main hypothesis under test. A good design is one that effectively controls for (many) possible sources of error."

Definition 4.1 (Controlling Observables in Experimentation): *When an experimentalist holds observable variables constant or randomly assigns them to evaluate the effect of one or more treatments on subjects' choices.*

Definition 4.2 (Script): *The context of the instructions and information given to subjects in an experiment.*

4.1.2 Controlling Unobservables in Laboratory Experiments

Control can also mitigate problems from subject-specific unobservable variables when a laboratory researcher uses a within-subjects design as discussed in Section 3.3.3. That is, by using a within-subjects design a researcher can hold constant things about the subject that are unobservable such as interest in the experiment, overall mood, and cognitive ability.

Sometimes laboratory experiments can make some variables observable that are typically unobservable without experimental manipulation and, thus, enable a researcher to control these typically unobservable variables, as discussed in Chapter 2. For example, in political economy laboratory experiments, as we saw in the Battaglini, Morton, and Palfrey experiment (Example 2.6), the researchers use financial incentives to motivate subjects to take their choices in the experiment seriously, to make the choices salient to the subjects. Holding these financial incentives constant, Battaglini, Morton, and Palfrey then manipulated other aspects of the experimental environment. Thus, Battaglini, Morton, and Palfrey control subjects' motivations to some extent.

Why might financial incentives help control unobservables? Suppose that we suspect that voters who have more intense preferences for candidates are more likely to be informed and more likely to vote, but there is no way to accurately measure variation in voter preference intensity in observational data. Without being able to control for intensity, it is possible that this unobservable is confounding the observed relationship between information and voting. In a laboratory or web-based election, voters can be paid based on the outcome of the election, and the size of the payoff can be set to control for intensity effects. (That is, the researcher can hold voter payoffs according to preference orderings equal across voters such that each voter receives the same payoff if her first preference wins, and so forth.) In this fashion, voter intensity can be held constant across voters. Of course, this raises other issues about the comparability of such experiments to voter intensities in observational data. Nevertheless, many of the measures used in laboratory and virtual laboratory experiments on the web are used to control both observable variables and, in particular, unobservable variables outside the laboratory. Through control, then, the researcher can more safely calculate treatment effects than with observational data.

Definition 4.3 (Controlling Unobservables in Experimentation): *When an experimentalist attempts to control typical unobservables through within-subjects designs, by manipulation, or by observation to evaluate the effect of one or more treatments on subjects' choices.*

Another example of control over a normally unobservable variable is how much time and effort individuals spend on a task. In a laboratory experiment, researchers can manage how subjects spend their time on various tasks and actually measure how much time subjects spend on one task instead

of another, whereas outside of the laboratory, researchers cannot typically observe how subjects or individuals in general allocate their time to various tasks. We discuss Example 4.2 later in this chapter, in which researchers both control and monitor the time that subjects spend on various pieces of information during a laboratory election campaign.

Finally, we present an especially interesting method that political psychologists have used to measure racial attitudes through the use of subliminal primes (words displayed to subjects that are viewed unconsciously) coupled with implicit measures of responses in Example 4.1. In a set of experiments, Taber (2009) evaluates the theory that racism and prejudice are no longer significant reasons why individuals object to policies such as affirmative action and that instead conservative principles such as individualism and opposition to big government explain such objections. However, measuring racial prejudice is extremely difficult observationally or in simple surveys given the stigma attached to such preferences. In one of the experiments he conducts, Taber exposes subjects to the subliminal prime of affirmative action and then measures the time it takes for them to identify words related to racial stereotypes, conservative principles, and a baseline manipulation of unrelated words. The subjects are told that their job is to identify words versus nonwords, and they are exposed to nonwords as well.

Example 4.1 (Subliminal Priming Lab Experiment): *Taber (2009) conducted a series of experiments in which he measured the effects of subliminal primes of the words affirmative action and welfare on implicit responses to racial and gender stereotypes and conservative individualist principles.*

Target Population and Sample: *Taber used 1,082 voting-age adults from five U.S. cities (Portland, Oregon: 90; Johnson City, Tennessee: 372; Nashville, Tennessee: 132; Peoria, Illinois: 138; and Chicago, Illinois: 350). The subjects were recruited by print and Internet advertisements in the summer of 2007. “The sample included: 590 men, 492 women; 604 whites, 364 blacks, 104 other; 220 self-reported conservatives, 488 liberals, 332 moderates; 468 with household income below \$15,000, 260 with income \$15,000–30,000, 354 with income greater than \$30,000; 806 with less than a college diploma. The mean age was 40 years with a range of 18 to 85 years.”*

Subject Compensation: *Subjects were paid \$20 for participating.*

Environment: *“Participants came to an experimental location at an appointed time in groups of no more than eight. Laptop computers were set up in hotel or public library conference rooms in a configuration designed to minimize distractions. The . . . experiments were programmed in the MediaLab and DirectRT software environment and run on identical Dell laptop computers,*

proceeded in fixed order, with the pace controlled by the participant. All instructions appeared onscreen. Participants were consented before the session, debriefed and paid \$20 after.” We discuss the benefits of debriefing in Sections 12.1.2 and 13.6.3.

Procedures: *The subjects participated in six consecutive experiments in a single, one-hour session. Subjects were also given a survey of political attitudes, demographics, and so on. We describe each experiment in the order in which it was conducted.*

Study 1: *Subjects were first given a subliminal prime of the phrase “affirmative action” and then a target word or nonword, which the subject was asked to identify as either a word or nonword. The target words came from six sets of words with an equal number of nonword foils. The nonwords were pronounceable anagrams. The six sets were (p. 10) “Black stereotype targets (rhythm, hip-hop, basketball, hostile, gang, nigger); White stereotype targets (educated, hopeful, ambitious, weak, greedy, uptight); female stereotype targets (caring, nurturing, sociable, gossipy, jealous, fickle); individualism targets (earn, work-ethic, merit, unfair, undeserved, hand-outs); egalitarianism targets (equality, opportunity, help, need, oppression, disadvantage); big government targets (government, public, Washington, bureaucracy, debt, mandate); and pure affect targets (gift, laughter, rainbow, death, demon, rabies). . . . In addition to these affirmative action trials, there were also interspersed an approximately equal number of trials involving the prime ‘immigration’ and a different set of targets” which Taber (2009) does not discuss. “In total, there were 72 affirmative action/real target trials, 72 baseline/real target trials, and 144 non-word trials, not including the immigration trials. On average study 1 took approximately ten minutes to complete.”*

Note that the target words were of three types: stereotype targets, principle targets, or baseline targets.

The prime and target were presented (pp. 7–8) “in the following way . . . : a forward mask of jumbled letters flashed center screen (e.g., KQHYTPDQFPBYL) for 13 ms, followed by a prime (e.g. affirmative action) for 39 ms, a backward mask (e.g. DQFPBYLKQHYTP) for 13 ms, and then a target (e.g., merit or retim, rhythm or myhrth), which remained on screen until the subject pressed a green (Yes, a word) or red (No, not a word) button. Trials were separated by a one second interval. Where precise timing is critical, masks are necessary to standardize (i.e., overwrite) the contents of visual memory and to ensure that the effective presentation of the prime is actually just 39 ms. Conscious expectancies require around 300 ms to develop.”

Taber measured the response times on word trials, discarding the nonword trials.

Study 2: Subjects “were asked to think about affirmative action and told that they might be asked to discuss this issue with another participant after the study. One third were told that this discussion partner would be a conservative opponent of affirmative action, one third were told to expect a liberal supporter of affirmative action, and for one third the discussion partner was left unspecified.” Then subjects completed the same task of identifying words and nonwords as in study 1 without the subliminal primes.

Study 3: In this study Taber used the race stereotype words as primes for the principle targets and vice versa, mixed in with a larger set of trials designed to test unreported hypotheses. He used the same procedure in the subliminal primes as in study 1.

Study 4: This study used black and white stereotype words as primes for pure affect target words using an equal number of positive and negative examples.

Study 5: Taber conducted a replication of a famous experiment conducted by Sniderman and Carmines (1997). “Participants . . . read a realistic one-page description of a fictional school funding proposal that sought to provide \$30 to \$60 million per year to disadvantaged school districts in the participant’s home state. The proposal was broken into an initial summary, which manipulated whether the program would be publicly or privately funded, and a brief case study of a particular school that would receive funding through the proposed program, which manipulated race of recipients in three conditions . . . : the school was described as predominantly white, black, or racially mixed. . . . After reading the summary and case study, participants were asked a single question . . . : Do you support of [sic] oppose this proposed policy? Responses were collected on a 7 pt. Likert-type scale” (p. 21).

Study 6: This study replicated study 5 “with a simpler affirmative action program using different manipulations.” Taber manipulates “need versus merit, and target race, but this time the brief proposal mentions a particular disadvantaged child as a target recipient. The race of the child is subtly manipulated by using stereotypical white, black and racially-ambiguous names (Brandon, Jamar, and James, respectively). The child is described either as struggling academically with a need for special tutoring he cannot afford or as a high achieving student who would be targeted by the program because of exceptional ability and effort” (p. 23). Subjects were asked again whether they support or oppose the proposed program on a seven-point scale.

Results: In study 1, Taber found evidence that both conservative opponents of affirmative action and liberal supporters of affirmative action had shorter response times to black stereotype targets as compared to other targets. In study 2, he found that explicitly thinking about affirmative action in the absence of an expectation about the discussion partner led to shorter response times to

principle and race stereotype targets. When conservative opponents expect to talk to a like-minded partner, only response times for black stereotype targets are shorter; when they expect to talk to a liberal supporter, black stereotype target response times are shorter than the baseline but also are principle target response times. Liberal supporters’ response times are shorter for black and principle targets regardless of whom they expect as a discussion partner. In study 3, Taber found that race primes reduced reaction times on principle targets, particularly for conservatives.

Taber used the data from study 4 plus survey data on the subjects to devise a measure of implicit affect toward African Americans. He found that white respondents had more negative implicit attitudes toward African Americans than black respondents and that conservative opponents of affirmative action were significantly more negative toward blacks than white supporters. Taber’s analysis of the data from studies 5 and 6 supports previous results found by Sniderman and Carmines and Feldman and Huddy (2005) that conservatives prefer private to public funding, white liberals prefer public funding, and neither group has a significant preference for racial targets. However, he finds that conservatives who are politically sophisticated strongly oppose spending public funds when it targets black schools (students) but not when white or mixed race schools (students) are targeted.

Comments: Taber’s studies are a good example of building on previous research in psychology on the subliminal prime effects on response times. By using the primes and measuring response times to particular types of words that have known connotations, he was able to make observable how subjects respond to these words. In studies 1–4 he made use of a within-subjects design, whereas in studies 5 and 6 he used a between-subjects design (see Section 3.3.3).

Taber found that response times for words related to racial stereotypes are affected by the subliminal primes for conservative opponents of affirmative action as compared to other words, which he argues suggests that racial prejudice is a factor in explaining conservative opposition to the policy. Observing such response times to subliminal primes would be difficult to accomplish outside of a laboratory environment. That said, Taber took the lab to his subjects to some extent, recruiting the subjects to temporary laboratories set up in local library or hotel conference rooms in five U.S. cities. Such an experiment is what is often called a “lab in the field” experiment, which we investigate more fully in Section 8.2.3 and define in Definition 8.5.

Taber’s experiment is also a good example of how a researcher working with naturally occurring words and situations may need to conduct a manipulation check to be sure that the manipulation he or she is conducting

captures the manipulation he or she wishes to conduct. He used as his primes words that had been shown in previous psychological studies to fit the categories of interest. He also checked that the names he assigned to the child in study 6 implied a particular race for the child; he found that 73% of the participants in the Brandon Smith condition perceived him as white, 88% perceived Jamar Smith as black, and 64% were uncertain about James Smith.

Definition 4.4 (Manipulation Check): *A survey or other method used to check whether the manipulation conducted in an experiment is perceived by the subjects as the experimenter wishes it to be perceived.*

The same degree of control is not generally possible when conducting field experiments. First, it is not generally possible to gather repeated observations on the same subject and control for unobservables in this fashion when an experiment is conducted in the field, although it may be possible via the Internet. While in the laboratory or via the web, a researcher can induce preference orderings over candidates; in field experiments, researchers investigating the effect of information on voting must work within the context of a given election that he or she cannot control or of a set of elections and the unobservable aspects of voter preferences in those elections. Hence, researchers using field experiments focus more on how random assignment can help determine causality rather than the combination of control and random assignment, whereas researchers using laboratory experiments (both physical and virtual) use both control and random assignment in designing experiments. Unfortunately, random assignment is harder to implement in the field as well because experimenters confront problems of nonresponse and noncompliance, so in many cases field experimentalists must often also rely on the statistical methods discussed earlier to deal with these problems, and these statistical methods require making untestable assumptions, as we show below.

4.2 Control Functions in Regressions

4.2.1 When an Observable Variable Cannot Be Controlled or Manipulated

What happens when a researcher is investigating the effects of information on voting behavior in observational data or in some cases experimental data gathered through a field experiment in which the researcher did not